

METHODS

Data Collection

The MCR collects reports of all newly diagnosed cancer cases from all Massachusetts acute care hospitals and one health maintenance organization (83 reporting facilities in 1998). The MCR compiles summaries of cancer incidence, such as this report, and also produces special reports. These undertakings require data collection efforts that necessitate extensive interaction with hospital tumor registrars. Intensive data evaluation is also required to ensure data quality. The fundamental requirements of any central cancer registry include: (1) complete registration, (2) prevention of case duplication, (3) collection of uniform data, i.e., standardization of items, definitions, rules, classification and nomenclature of primary site, histology, staging and procedures, (4) quality control, and (5) efficient data processing.

The data summarized in this report are drawn from data entered on MCR computer files on or before May 4, 2001. The numbers herein may change slightly in future reports, reflecting late reported cases, address corrections, or other changes based on subsequent details from reporting facilities. Furthermore, as health researchers may use these data to meet a diverse range of needs, they may produce results slightly different from those published herein. Using slightly different population estimates or statistical methodologies, such as grouping ages differently or rounding off numbers at different points during calculations, may also produce results slightly different from those published in this report.

MCR case ascertainment improved during the years covered by this report. For diagnosis year 1998, the MCR's total case count was estimated (by the North American Association of Central Cancer Registries) to be complete. This report includes (for the diagnosis years indicated) two case sources that were not available for most previous editions of the *City and Town Supplement* -- physician office cases and death certificate-only cases:

For diagnosis years 1996 and onward, the MCR collects information from reporting hospitals, where available, on cases diagnosed and treated in staff physician offices. Not all hospitals report this type of case, however, and some hospitals report such cases as if the patients had been diagnosed and treated by the hospital directly. Collecting this type of data makes the MCR's overall case ascertainment more complete, but because these cases are not reported by every hospital, there will be effects on the reporting completeness in some geographic areas. If a certain hospital reports physician office cases to the MCR and mainly serves patients living in one geographic region of the state, for example, the case collection of that region's cancers may be slightly more complete than that of other regions where hospitals do not report physician office cases.

For diagnosis years 1997 and 1998, the MCR identified previously unreported cancer cases through death certificate clearance to further improve case completeness. In some instances, a cancer-related cause of death recorded on a Massachusetts death certificate is the *only* source of information for a cancer case. These "death certificate-only" cancer diagnoses are therefore poorly documented and have not been medically confirmed (confirmed by review of complete clinical information). Such cases are included in this report for diagnosis years 1997 and 1998.¹

Coding for cancer types in this report follows the International Classification of Diseases for Oncology (Second Edition) system (see **APPENDIX D**). The list of reportable neoplasms is the same as that used for the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program data, with the exception of *in situ* neoplasms. The MCR began collecting information on *in situ* neoplasms diagnosed as of January 1, 1992; however, *in situ* cases are **not** included in this report. You may contact the MCR for information on *in situ* neoplasms.

¹ The death certificate review process also identifies some cancer cases that are found to have been diagnosed before death but are not reportable to the MCR; such cases are not added to MCR databases and are not included in this report. The process also identifies some cancer cases that were diagnosed before death and should have been reported to the MCR; these previously "missed" cases are added to MCR data files for the appropriate diagnosis years and are included in this report for 1994-1998.

Data Presentation

Three measures of cancer incidence are presented in this report's data tables: expected case counts, observed case counts, and standardized incidence ratios (SIRs).

Expected and Observed Case Counts

In this report, the *observed* case count (**Obs**) for a particular type of cancer in a city/town is the actual number of newly diagnosed cases reported to have been diagnosed in residents of that city/town from 1994 through 1998. The "Total" observed case count for each cancer type is the sum of the number of observed male and female cases only. The MCR added two additional gender classifications (transsexuals and persons with sex chromosome abnormalities/hermaphrodites) for cases diagnosed as of January 1, 1995. (Cases diagnosed before this date were limited to male or female only.) Any case classified in either of the new gender categories² is *not* included in this report because the population data used in the statistical calculations only include male and female categories.

A city/town's *expected* case count (**Exp**) for a certain type of cancer for this time period is a calculated number based on that city/town's population distribution (by sex and among six age groups) for 1996 (the midpoint of 1994 through 1998), and the corresponding statewide average age-specific incidence rates. See ***Calculation of an SIR*** (below) for an example of how a hypothetical expected count is calculated. The expected case counts in this report are rounded to the nearest hundredth (two decimal places); if the total expected case count is not exactly equal to the sum of the male and female expected counts, this is attributable to rounding error.

For the computation of statewide average age-specific incidence rates used for this report, the 1996 statewide population estimates (by sex and six age groups) were obtained from the Massachusetts Institute for Social and Economic Research (MISER)³. Different methodologies may be used to derive slightly different population estimates, yielding slightly different results.

Standardized Incidence Ratios

The data tables present SIRs (rounded to the nearest whole number) for males, females and the total population of each city/town for twenty-three types of cancer and for all cancers combined. An SIR is an indirect method of adjustment for age and sex that describes in numerical terms how a city/town's average experience in 1994-1998 compared to that of the state as a whole. The SIR is a useful tool for screening incidence data and generating leads for further public health investigations.

- An SIR of *exactly 100* indicates that a city/town's incidence for a certain type of cancer is *equal to that expected* based on statewide average age-specific incidence rates.
- An SIR of *more than 100* indicates that a city/town's incidence for a certain type of cancer is *higher than expected* for that type of cancer based on statewide average age-specific incidence rates. For example, an SIR of 105 indicates that a city/town's cancer incidence is 5% higher than expected based on statewide average age-specific incidence rates.
- An SIR of *less than 100* indicates that a city/town's incidence for a certain type of cancer is *lower than expected* based on statewide average age-specific incidence rates. For example, an SIR of 85 indicates that

² Nineteen cases classified in the new gender categories are recorded at the MCR for 1994-1998.

³ The MISER population estimates for 1996 were released in November 1999.

a city/town's cancer incidence is 15% lower than expected based on statewide average age-specific incidence rates.

Measures of Statistical Significance

Tests of statistical significance allow an estimate of the probability that the difference between the observed and expected case count is due to chance alone. This estimate is referred to as a "**p**" value. A **p** value of less than or equal to 5% ($p \leq 0.05$) means that there is, at most, a 5% chance that the difference between the observed and expected case count is due to chance alone; thus, a cancer excess or deficit with such a **p** value is considered statistically significant. *The presence or absence of statistical significance does not necessarily imply biological or public health significance.*

In this report, three levels of statistical significance are employed to identify cities and towns with excess cancer incidence (and deficits) as compared with statewide average incidence -- $p \leq 0.05$, $p \leq 0.01$, and $p \leq 0.001$. The use of $p \leq 0.001$ highlights those cancer excesses least likely to have occurred by chance alone. Use of this stringent criteria, however, makes it difficult to identify elevated SIRs for towns with relatively small populations and small numbers of cancer cases. The use of $p \leq 0.05$ constitutes a less stringent criterion and identifies a greater number of cancer excesses. Use of $p \leq 0.05$ can provide investigators with a broader context for identifying patterns of excess cancer incidence than use of $p \leq 0.01$ or $p \leq 0.001$.

$p \leq 0.05$: In the data tables, $p \leq 0.05$ is used to identify cancer types having significant excesses or deficits at the least stringent level used herein -- $p \leq 0.05$. This indicates that there is, at most, 1 chance in 20 that the identified excess or deficit of cancer cases is due to chance alone. A pound symbol (#) following an SIR marks that excess or deficit as being statistically significant at the $p \leq 0.05$ level, but not at the higher levels ($p \leq 0.01$ and $p \leq 0.001$). Based on the number of tests performed for this report (eighteen male/female sites and five single-sex sites), one would expect by chance alone to find 360 significant excesses at the $p \leq 0.05$ level; 393 were found.

$p \leq 0.01$: A **p** value of less than or equal to 0.01 indicates that there is, at most, 1 chance in 100 that the excess or deficit of cancer cases is due to chance alone. (Note that all cancer excesses and deficits which are statistically significant at this level are also significant at the less stringent $p \leq 0.05$ level, but not all data significant at the $p \leq 0.05$ level are significant at the $p \leq 0.01$ level.) A tilde symbol (~) following an SIR indicates that these data are significant at both the $p \leq 0.05$ and $p \leq 0.01$ levels, but not at the more stringent $p \leq 0.001$ level. Based on the number of tests performed for this report, one would expect by chance alone to find 72 significant excesses at the $p \leq 0.01$ level; 160 were found.

$p \leq 0.001$: This is the most stringent criterion employed in this report to highlight cancer excesses and deficits that are least likely to be due to chance alone. A **p** value of less than or equal to 0.001 means that there is, at most, 1 chance in 1000 that the excess or deficit in observed cases is due to chance alone. A caret symbol (^) following an SIR indicates that these data are significant at all three levels of significance testing used in this report. Based on the number of tests performed for this report, one would expect by chance alone to find 7 significant excesses at the $p \leq 0.001$ level; 52 were found.

Calculation of an SIR

$$\text{SIR} = (\text{OBSERVED} / \text{EXPECTED}) \times 100$$

The following example illustrates the method of calculation for a hypothetical town for one type of cancer for the year 1996:

	<u>Town X</u>	<u>State</u>	<u>Town X</u>	<u>Town X</u>
Age Group	Population (A)	Age-Specific Incidence Rate (B)	Expected Cases (C) = (A) x (B)	Observed Cases (D)
0-19	74,657	0.0001	7.47	11
20-44	134,957	0.0002	26.99	25
45-64	54,463	0.0005	27.23	30
65-74	25,136	0.0015	37.70	40
75-84	17,012	0.0018	30.62	30
85+	6,337	0.0010	6.34	8
total:			136.35	144

$SIR = (\text{Observed Cases}/\text{Expected Cases}) \times 100 = (\text{column D total}/\text{column C total}) \times 100 = (144/136.35) \times 100 = 106$

Thus, Town X's incidence for this type of cancer is approximately 6% higher than the corresponding statewide average incidence for this type of cancer.

Data Limitations

It should be remembered that apparent increases or decreases in cancer incidence over time may reflect changes in diagnostic methods or case reporting rather than true changes in cancer incidence. Three other limitations must be considered when interpreting cancer incidence data for Massachusetts cities and towns: under-reporting in areas close to neighboring states; under-reporting for cancers that may not be diagnosed in hospitals; and cases being assigned to incorrect cities/towns.

Border Areas and Neighboring States

Some areas of Massachusetts appear to have low cancer incidence, but this may be the result of under-reporting - that is, a loss of cases diagnosed or treated in neighboring states that are not reported to the MCR. Presently the MCR has reciprocal reporting agreements with fifteen states -- Alaska, Arkansas, Connecticut, Florida, Maine, Mississippi, New Hampshire, New York, North Carolina, Rhode Island, South Carolina, Texas, Vermont, Wisconsin and Wyoming.

Cases Diagnosed in Non-Hospital Settings

During the time period covered by this report (1994 through 1998), the MCR's information sources for nearly all newly diagnosed cancer cases were hospitals. Some types of cancer in this report are undoubtedly under-reported because they may be diagnosed by private physicians, private laboratories, health maintenance organizations or radiotherapy centers that escape hospital case identification systems. Examples may include melanoma of skin, prostate cancer, and certain hematologic malignancies such as leukemia and multiple myeloma. The extent of this under-reporting has not been determined exactly, but cases included in this report represent the great majority of actual cases and provide an essential basis for observing cancer incidence patterns.

City/Town Misassignment

In accordance with standard central cancer registry procedures, each case reported to the MCR should ideally be assigned to the city/town in which the patient lived at the time of diagnosis, based on the patient address provided by the reporting hospital. In practice, however, a patient may provide the hospital with his/her mailing address (e.g., a post office box located outside the patient's city/town of residence); a business address; a temporary address (e.g., the patient is staying with a relative while receiving treatment and reports the relative's address as his/her own); or a locality or post office name (e.g., "Chestnut Hill" rather than "Boston", "Brookline" or "Newton"). In addition, if a patient has moved since being diagnosed, the hospital may report the patient's current address rather than that at the time of diagnosis. Because of the large number of cases reported to the MCR, and because most data are reported to the MCR via electronic media, most city/town case assignments are performed by an automated computer process. This simplified matching process may misassign some cases based on the reported locality name. When MCR staff become aware of such misassignments, the errors are corrected manually. Furthermore, in order to minimize such errors, cases from almost forty geographic localities prone to city/town misassignment are processed manually by the MCR.